

SYMBOLIC MODEL OF PERCEPTION IN DYNAMIC 3D ENVIRONMENTS

D. W. Carruth*, B. Robbins, M. D. Thomas, and A. Morais
Center for Advanced Vehicular Systems
Mississippi State, MS, 39762

M. Letherwood and K. Nebel
RDECOM/TARDEC
Warren, MI 48397

Approved for public release: distribution is unlimited

ABSTRACT

Computational models of human cognition have been applied to many complex real-world tasks including air traffic control, human-computer interaction, learning arithmetic, traversing the World Wide Web, intelligent tutors, instrument-based flight, and vehicle driving. There are numerous additional applications for these computational models including integration with models of human motion, military simulation of enemy agents in virtual environment training, testing of new vehicle designs or machine interfaces, and analysis of cognitive components of tasks. However, most of these models exist in limited two-dimensional (2D) environments. In order to apply computational models to tasks in a dynamic three-dimensional world, extensions to current cognitive architectures must provide the capability for models to perceive, process, and act in the three-dimensional environments. The current research will seek to extend the vision components of a cognitive architecture to support computational models capable of simulating human vision in a dynamic three-dimensional (3D) environment.

1. INTRODUCTION

The Future Combat Systems (FCS) consist of 18 systems including unattended ground sensors, intelligent munitions systems, 6 classes of unmanned vehicles, the multifunction utility/logistics and equipment vehicle, and 8 classes of warfighting or support manned vehicles. Each of these systems comes into direct or indirect contact with the soldier at many points introducing potential cognitive, ergonomic, and performance issues. The Future Force Warrior (FFW) soldier is treated as an integral part of the FCS and is expected to benefit from advanced technology in networking, computing, environment and physiological monitoring, and armor. The goal of the FFW system is to increase effectiveness and flexibility while decreasing load. In order to effectively build and deploy the FFW systems each of the components will need to be designed,

prototyped, lab tested, revised, field tested, and further revised until the system is operational and effective.

The researchers, designers and engineers involved in the development of the FCS and FFW systems will likely utilize computer-aided design (CAD), computer-aided engineering (CAE) and computational simulation tools such as finite element analysis (FEA) in order to reduce development time, testing time, and overall costs. CAD/CAE and FEA are computational tools that allow designers and engineers to design and test the material and mechanical properties of virtual prototypes. The benefits of these design capabilities are such that they enjoy widespread use in many major industries. These systems have little or no built-in support for determining the needs of the constraints imposed by the end user (Porter, Case, & Freer, 1999).

Existing design tools such as Jack, RAMSIS, and SAFEWORK provide limited abilities to evaluate human interaction with virtual prototypes. However, these models focus primarily on user attributes such as anthropometry, viewing volumes, and static postures. These models do not predict complex human task performance and interaction with virtual designs (Porter, Case, & Freer, 1999). Computational cognitive architectures may provide a partial solution to simulating the role of the soldier in FCS and the FFW systems. These software architectures provide a framework for cognitive scientists and human factors engineers to create models capable of simulating human task performance. Current models cannot simulate the entirety of human cognition from sensory input to mental processing to the execution of motor actions. However, a number of architectures (e.g. ACT-R, EPIC, MIDAS, SOAR, QN-MHP) have made it a goal to attempt to define a formal theory of human perception, cognition, and action. These architectures may be able to provide predictive capabilities for consideration of human interaction with FCS and the FFW system. These architectures have been applied to real-world tasks such as driving vehicles (Salvucci, 2006; Liu, Feyen & Tsimhoni, 2006) and piloting UAVs (Ball, Gluck, Krusmark, & Rodgers,

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 01 NOV 2006		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Symbolic Model Of Perception In Dynamic 3d Environments				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Center for Advanced Vehicular Systems Mississippi State, MS, 39762				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM002075.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 8	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

2003). However, current models have limited dynamic 3D perceptual and motor capabilities.

Rather than develop an entirely new architecture, the goal of the current research is to extend an existing cognitive architecture, ACT-R (Anderson, Bothell, Byrne, Douglass, Lebiere, and Qin, 2004). ACT-R has been used to simulate vision in dynamic 3D environments, most notably in Salvucci's model of the driver (2006). The extensions to ACT-R in that case were at least partly specific to the task of driving. The current project intends to place a digital human model within a virtual environment generated by a commercial-off-the-shelf (COTS) software package. In doing so, we hope to create a platform that will allow models of other real world tasks to use our more general extensions to the existing ACT-R architecture. This paper relates the development of extensions to the vision module of ACT-R. These extensions include modifications to motion perception and the encoding of spatial information.

The current paper provides an overview of the ACT-R cognitive architecture, some details about ACT-R's current vision module, and details regarding the design and implementation of each of our extensions.

2. ARCHITECTURE FOR MODELING

The development of models of human performance takes place within an architecture that combines a simulation of human cognition with an environment in which the digital human can perform. Our interest is in developing models of whole-body real-world tasks such as vehicle maintenance or driving. Such models require an extensive model of the environment and a capable, but general model of human cognition. The current research focused on extending the ACT-R cognitive architecture (Anderson, et al., 2004) and the Virtools virtual environment.

2.1. ACT-R Cognitive Architecture

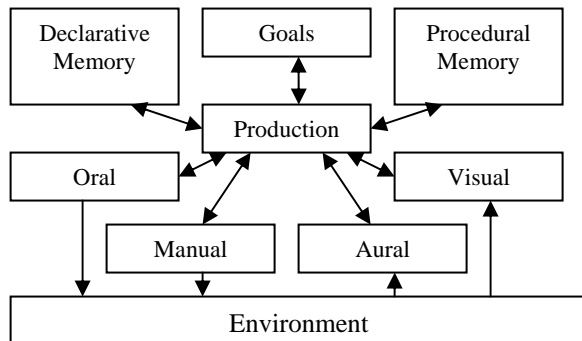


Fig. 1. Diagram of the ACT-R architecture.

ACT-R is a hybrid symbolic/subsymbolic cognitive architecture that allows designers to develop computational models of human performance (Anderson, et al., 2004). At its core, the ACT-R architecture (see Figure 1) is a production system architecture where procedural knowledge is represented by condition-action rules, known as productions. As the architecture executes a model, the condition of the production is tested against the model's awareness of the current state of the environment. A single production from the set of matching productions is selected based on expected utility and the production's actions are applied. A set of modular systems wrapped around the core production system attempt to execute the actions requested by the fired production.

The modular systems simulate declarative memory, goal tracking, vision, audition, and motor actions. The ACT-R memory system includes two basic types of memory. Procedural memory is the store of condition-action productions that are selected and executed by the core production system. Declarative memory is a store of memory units called *chunks* and simulates the storage, activation, and retrieval of memory. The goals system tracks the current goal of the cognitive model. The visual and aural systems encode sensory inputs from the environment as chunks of declarative knowledge. The oral and manual systems provide mechanisms by which the cognitive model can act on the environment. By posting module requests, the cognitive model can retrieve memories, work on goals, recognize visual or aural percepts, and perform actions.

ACT-R has been selected for the current research for three primary reasons. First, it has a wide range of capabilities in the existing cognition, perception, and action modules. The modular nature of the architecture allows us to extend the vision module while retaining the benefits of the existing modules. Second, ACT-R has been successfully applied to a variety of tasks from basic memory tasks to advanced tasks such as driving. Finally, ACT-R generates quantitative data, including reaction times, which can be directly compared to human performance data for validation purposes.

2.2. Virtual Environment

Virtools, a commercially available graphics rendering and environment development software package, is used to generate the virtual environment. Virtools allows environment developers to construct a 3D virtual environment consisting of modeled objects and scripts that define object behaviors. Objects modeled in the virtual environment are tagged with symbolic information such as the object category, text content, and any other visual properties that would be difficult or impossible to extract directly from the visual display.

A Virtools/ACT-R software interface tracks the location of the digital human model within the environment and renders the digital human model's 200° horizontal by 40° vertical view of the environment every 17 ms. The digital human model's visual field is processed to determine which objects in the virtual environment are currently visible. Intrinsic symbolic information, such as spatial location and size in visual angle, is calculated and the extrinsic symbolic information associated with the objects by the modelers is extracted from the rendered visual field and the environment. This symbolic information is then passed over a network connection to a running ACT-R model that updates the internal representations of the vision module with the new information.

The vision module is a symbolic model of human visual perception in which the symbolic information that can be encoded from each object is directly supplied by the environment. While machine vision research is progressing, it would be computationally impractical to process video from real or virtual environments. Instead, the object attributes stored within the virtual environment provide the end results of the vision process without requiring that we simulate the entire visual process. This symbolic model of vision allows us to focus on modeling the complex interactions between attention, cognition, and action without being overly concerned about the details of sensation and perception.

3. SYMBOLIC MODEL OF VISION

ACT-R's vision module is based on a feature theory of perception that synthesizes multiple existing theories of visual perception (feature-integration theory; Triesman & Gelade, 1980), attention (attentional spotlight; Posner, 1980) and search (guided search; Wolfe, 1994). The vision module's representation of the visual field is a visual icon that contains all of the features that are currently visible in the 200° by 40° visual field rendered and processed by the virtual environment. An ACT-R model cannot directly access the features stored in the visual icon. The features must be the focus of visual attention in order to be extracted from the visual icon as declarative memory chunks representing the objects in the visual field. Two separate systems (visual-location and visual-object) provide the mechanisms that allow ACT-R to extract these chunks.

The visual-location system implements a preattentive search for features and conjunctions of features. The 2D locations of visual features are always available in ACT-R. When the model requests a preattentive search for a visual location, the model must specify a set of constraints that will guide the search of the visual field. The visual-location system will immediately return a single location

that matches the specified constraints. The visual-location systems allow constraints on visual properties (color, motion, size), spatial location, and whether previously attended (Anderson, et al., 2004). By allowing multiple constraints to be specified, ACT-R allows largely unconstrained searches for conjunctions of features similar to Wolfe's guided search (Wolfe, 1994). The basic visual properties supported by ACT-R include color, size, and type. Modelers can extend the visual icon to include additional features that can be used to constrain searches.

Spatial location constraints allow ACT-R to globally limit search to spatial areas of the visual field or to constrain search relative to the current focus of attention. For example, if the current goal is to monitor a red indicator light that is known to appear at the top left of the display, the visual-location search can be constrained to return the location of red features in the top left of the visual field. An example of a relative constraint would be to search for the nearest matching location to the current focus of attention.

ACT-R tracks whether objects in the visual field have been attended or not. Each object is flagged with one of three states: recently attended, not recently attended, and recently onset. At onset, the newest object may be pushed into the visual-object system forcing attention to shift to the object by what is referred to as buffer stuffing. Buffer stuffing provides a limited simulation of bottom-up attentional capture.

When the core production system specifies a set of constraints, the visual-location system determines whether any location in the visual field matches the constraints. If not, the vision module is set into an error state that must be addressed by the core production system. If one location matches the constraints, the preattentive search generates pop-out effects (Triesman & Gelade, 1980). If more than one location matches, one of the locations is randomly selected and returned. In order to find the desired target, a self-terminating serial search must be performed by the vision module.

The visual-location system provides a location matching some basic constraints but, in order to recognize the features as an object, the attentional spotlight must be shifted to the location. By default, ACT-R models attentional shifts and not the actual movement of the eyes (Anderson, et al., 2004; Anderson, Matessa, & Lebiere, 1998). Salvucci's EMMA (2001) extension to ACT-R allows modelers to simulate eye movements. Once attention is shifted to a location, object features are bound and encoded into a declarative memory chunk that is made available to the core production system through the vision module.

The vision module has been successfully used in computational models applied to a range of simple and complex tasks. However, it is missing certain capabilities that are necessary for simulating vision in dynamic 3D environments. These missing features include aspects of motion perception, 3D spatial perception, and coordination of shifts of attention with movements of the head.

3.1. Motion Perception

Motion is a basic features of visual perception that can guide attention (Wolfe, 1998; 1994). Motion is not included as a feature in the ACT-R vision module. In order to model visual attention in a dynamic 3D environment, we needed to extend the vision module to support motion as a feature.

3.1.1. Preattentive Search

Moving objects are particularly important from a visual perception perspective. McLeod, Driver, and Crisp (1988) reported that moving items may be efficiently located amongst stationary and/or moving objects. However, searching for stationary objects amongst moving objects is inefficient. Searching for an object that may be moving or stationary is also inefficient suggesting that search cannot be directed at stationary and moving objects at the same time.

The particular aspects of motion that constrain the preattentive search by the visual-location system are debatable. Some evidence has identified separate systems for encoding the direction and magnitude of motion and using either system for guiding preattentive search for moving objects (Driver, McLeod, & Dienes, 1992). However, the evidence is complicated and it appears that motion is composed of one or two features.

Regardless of the number of features, motion features appear to be available for preattentive search. Our extension to the ACT-R vision module introduces motion across the 2D visual field as a feature in the visual icon. When the visual icon is updated from the display information provided by the virtual environment, the motion of each object is calculated by ACT-R and placed into the visual icon as motion features. The extended vision module represents motion as two features: motion magnitude and motion direction. As future research clarifies the representation of motion detection, the implementation may be revisited.

Both motion magnitude and motion direction are calculated based on the displacement of the center of the object in the visual field from the previous frame to the current frame. This limits ACT-R to an instantaneous estimate of motion magnitude and direction. Motion

magnitude is defined as the displacement of the object's location in the visual field in degrees of visual angle per second. Motion direction is specified in degrees with 0° representing motion along the positive X axis (to the right in the visual field). For example, an object moving from bottom left to top right across the visual field at 100°/sec would be represented as an object with a motion-direction feature of 45° and a motion-magnitude feature of 100°/sec.

There are possible concerns with our implementation of motion perception and preattentive search. First, there is no search asymmetry in visual search for motion. As previously mentioned, search for a moving object among stationary objects is efficient. However, search for a stationary object among moving objects is not efficient (Wolfe, 1998). ACT-R does not currently implement mechanisms for search asymmetry. Second, while motion speed can guide search for a fast moving target among slower moving targets, our implementation's use (and ACT-R's use) of quantitative values for specifying constraints seems overly powerful. For both of these issues, Wolfe's guided search (1994) may provide a potential solution. Guided search hypothesizes that search constraints are specified using broadly-tuned channels rather than the quantitative values used in ACT-R. The ACT-R community has expressed interest in moving the vision module towards guided search (Anderson, et al., 2004) and modifying search constraints to be more broadly defined may be a good intermediate step toward that goal.

3.1.2. Attentional Capture and Motion

The preattentive processes exist to guide attention to interesting objects in the visual field (Wolfe, 1998). Visual attention is guided by bottom-up and top-down processes. Bottom-up guidance of attention is based purely on the salience of the features of the object. If an object's feature salience is particularly high, attention will be captured and drawn to the object. Top-down guidance is the deployment of attention driven by task-related expectations.

Visual attention is captured when a visual feature attracts attention, even when the feature is irrelevant to the current task. The abrupt onset of new items appears to be the strongest stimulus leading to attentional capture (Wolfe, 1998). Motion is also a particularly powerful visual feature (Wolfe, 1998; 1994) and we must consider what conditions will lead to attentional capture by motion features.

Motion does not capture attention (Hillstrom & Yantis, 1994) but the onset of motion may briefly capture attention. However, even the onset of motion does not always capture attention. Von Mühlenen, Rempel, and

Enns (2005) propose that attention is captured not by the onset of new items but rather attention is captured by unique spatial and temporal events. In this case, the onset of motion may strongly attract attention when the rest of the display is static. If the onset of motion occurs at the same time as the sudden onset of other objects, the newly appearing objects should more strongly attract attention.

In the buffer-stuffing implementation of attentional capture, whenever new objects enter the visual field, one of the objects may capture attention. The attention-capturing object's features are immediately encoded into a declarative memory object and pushed into the vision module's buffer effectively forcing attention to shift to the new object. This limited system works relatively well for the appearance of completely new objects in unchanging scenes.

With the addition of motion to scenes, two features associated with the moving objects (location and size) are often changing. These changes lead the vision module to determine that the object is new and worthy of attention. This may force attention to continually be drawn to the moving object. The intent of ACT-R's implementation of bottom-up attentional capture is that the onset of new objects should attract and capture attention. Change in location and size of an object due to continuous motion should not lead to attentional capture. Only the abrupt onset of new objects or abrupt changes in features should capture attention. In our extension to the vision module, the visual icon is able to identify when a new object appears and when a feature of an existing object changes.

It appears that for many feature changes, the changing object attracts attention. Changes in color have been shown to capture attention in certain cases (von Mühlenen, et al., 2005). Changes in color, like the onset of motion, are not as likely to capture attention as the abrupt onset of new items and, in fact, may not draw attention when occurring along with new items (von Mühlenen, et al., 2005). In ACT-R, the onset of several objects coinciding with the onset of motion or changes in features leads to multiple objects being marked as new objects worthy of attention. Only one of these objects can attract attention, leading to the onset of motion being missed when there are other new objects. For our purposes, this result is an acceptable approximation of capture by unique events (von Mühlenen, et al., 2005). A more thorough model would include an estimate of the saliency of feature changes; ranking the abrupt onset of new objects higher than the onset of motion and other feature changes in existing objects. Further investigation into attentional capture, especially capture related to motion, will guide future revisions of the vision module.

3.1.3. Encoding Motion

The ultimate purpose of the vision module is to model the binding of features into objects through the attentional spotlight mechanism. Our extensions to ACT-R's vision module have implemented a model of the perception of motion across the 2D visual field represented by the visual icon. It is not clear that the model should encode the quantitative features of 2D motion across the visual field in declarative memory. We are aware of no compelling reason to not make the motion magnitude and direction available to the core production system.

At the same time, it is clear that models of active observers must encode some spatial information in a 3D environment and, additionally, must deal with the movement of the observer. As the observer moves, the 2D movement of the objects will be updated but the 3D spatial movement of the objects in the world will only be updated if the object is moving relative to the world. This will allow the model to recognize motion resulting from self-motion and motion resulting from actual motion of objects in the environment. The next section describes the details of our extensions to support the extraction and use of spatial information from the environment.

3.2. Spatial Information

In order to interact with a 3D environment, an observer must be aware of the spatial arrangement of the environment. Tversky (2003) and Tversky, Morrison, Franklin, and Bryant (1999) describe three major spaces of spatial cognition: the space for navigation, the space around the body, and the space of the body. Each space is essential for full interaction with the environment. The representation of the space for navigation contains landmarks and paths that define a simplified 2D, map-like view of the environment. The space around the body is a 3D reference frame in which the location of objects is verbally described relative to three axes defined by the body: head/feet, front/back, and left/right. The space of the body is a proprioceptive and kinesthetic sense of where the parts of the body are and how they are moving. The current work is focused primarily on the space around the body – 3D relationships between the observer and the nearby environment.

As attention shifts around the environment, the spatial relation between the observer and the object that is currently the focus of attention is encoded into declarative memory. The spatial relationship includes the egocentric bearing and the egocentric distance of the observer to the object. The spatial representation of the attended object may be elaborated by encoding relationships between the currently attended object and another object, most likely a landmark (Shelton & McNamara, 2001).

There has recently been significant work related to spatial systems in ACT-R. Gunzelmann and Anderson (2006; Gunzelmann, Anderson, & Douglass, 2004) examined strategies for computing the correspondence between an egocentric view of a task and an allocentric view of a task. In their ACT-R implementation, spatial information from the egocentric view was encoded in two steps. In the first step the general egocentric location of the target was encoded. In the second step, the spatial information necessary to differentiate the model from the nearby objects was extracted from the display. More objects lead to more complex descriptions which, in turn, lead to slower performance. The representations generated by the second step may be verbal descriptions similar to those described by Tversky (2003) or mental images. Work with mental images led to the implementation of an imaginal buffer in ACT-R (Gunzelmann & Anderson, 2006) for modeling mental manipulations of images and a visuospatial working memory for visualizing spatial problems (Lyon, Gluck, & Gunzelmann, 2006).

Johnson, Wang, and Zhang (2003) implemented an ACT-R model that encoded not only the egocentric relationship between the observer and the object, but it also encoded the object-to-object relationship between the current focus of attention and the previous focus of attention. This provided a rich representation of the environment that is not tied to the observer's location and may be used to assist in identifying the landmarks, paths, and nodes that make up the mental representations of Tversky's (2003) space of navigation.

ACT-R/S is an implementation of a separate spatial module within ACT-R that encodes spatial information in egocentric representations that are continuously updated (Hiatt, Trafton, Harrison, & Schultz, 2006; Harrison & Schunn, 2003). ACT-R/S includes a configural system that encodes and updates mental representations of the space around the body and the space for navigation and a manipulative system that encodes information about objects for manipulation by the motor system.

While our model of encoding spatial information has many similar aspects to each of the existing models, our model also has significant differences. Our model is similar to ACT-R/S, but, instead of implementing a separate spatial module, we currently use declarative memory to store spatial information. We also implement object-to-object relationship encoding similar to Johnson, Wang and Zhang (2003) and do not implement automatic updating of egocentric spatial relations. At any given moment, the model has a limited awareness of the spatial relationships. If attention shifts to an object that is outside of the observer's field of view, the model can compute the egocentric spatial relationship between the observer and the object based on the stored object-to-object

relationships encoded through visual exploration of the space around the body.

3.2.1. Visual Search

In Wolfe's (1998) review of features that can be used for efficient search, a few cues appear to allow efficient guided search of objects arranged in three dimensions including shading, occlusion, texture cues, shadows, and stereoscopic depth (Wolfe, 1998). However, none of these cues are necessarily associated with the egocentric spatial relationships or observer motion cues that separate the 3D spatial representation from a 2D visual field representation + depth planes.

Royden, Wolfe, and Klemmen (2001) investigated whether optic flow was treated differently than other structured fields of distractors to allow efficient search. The results did not show that search for a stationary object was more efficient in an optic flow condition than in another structured moving field. In the existing ACT-R models of spatial representation (Gunzelmann, & Anderson, 2006; Hiatt, et al., 2006; Johnson, et al., 2003), attention is required to encode an egocentric relationship between an object and the observer. While individual depth cues may lead to efficient search, 3D spatial information does not appear to.

In our spatial extensions to ACT-R, we assume that the egocentric and object-to-object relationships extracted from the environment are not available for efficient, preattentive search and are only available after attention has been focused on the object.

3.2.2. Encoding

Our model is similar to the Johnson, Wang and Zhang (2003) model of object location in that we encode the egocentric relationship between the observer and the object when attention is focused on the object. Object-to-object relationships are also encoded for objects near the point of gaze. Objects that will serve as good landmarks (Shelton & McNamara, 2001) should be preferred for object-to-object relationships.

The egocentric spatial relationships encode the observer's bearing to, and distance from, the edge of the attended object. In addition to the egocentric bearing and distance, the spatial system also encodes the direction and magnitude of the object in 3D space. The direction and magnitude is calculated based on the change in the egocentric relationship during the encoding time.

The bearing and distance from the previously attended object to the currently attended object may also be encoded as in Johnson, Wang and Zhang (2003). These object-to-object relationships are secondary to the

egocentric relationships but are essential for building spatial memory of the space for navigation (Shelton & McNamara, 2001). Some objects may be identified as landmarks or as providing special spatial relationships (e.g. walls, portals, readily visible features) and may be used to build hierarchical frames of reference.

Egocentric representations require regular updating as the observer moves through the environment (Tversky, 2003). Rather than continuously updating based on motor cues or a visual mechanism (i.e. optic flow), the model updates only the egocentric relationship and object-to-object relationships of those objects currently in the field of view. During motion, the model covertly and overtly shifts attention to objects in the environment to maintain the model's current awareness of the environment. The updating of the mental representation of spatial relations may not be automatic (Waller, Montello, Richardson, & Hegarty, 2002). Our implementation requires that the observer attend to the environment in order to update the mental representations of the spatial relations in the environment.

Spatial awareness of the environment provides the model with the capability to interact with 3D environments. The model can maintain awareness of objects and visual features that move in and out of the visual field as the observer moves through the 3D virtual space. The model can encode and update the 3D spatial location of objects and if the model needs to view an object outside of the current field of view, the model can request a rotation of the head to a remembered spatial location. The model is also able to request motor movements to spatial locations relative to the body. These motor movements allow the model to interact with objects in the 3D environment.

The spatial system is the most recent aspect of our extensions to be implemented and will require significant future work including support for imagining observer motion, memory for complex motion paths, building representations of space of navigation, and more.

3.3. Validation

Our extensions to the ACT-R vision module have not been validated against human data. The extensions to the vision module for motion perception largely mirror the implementation of perception for the other features ACT-R supports. The extensions for encoding spatial relationships may be more controversial and will require more effort to validate the performance of the model. Eye tracking and motion capture data on human performance in real-world, natural tasks such as model assembly, navigation and human-machine interaction is currently being captured in our lab (see these proceedings: Thomas, Carruth, McGinley, & Follett, 2006). These tasks will be

modeled using our extensions to the vision module and the results will be quantitatively compared to human data for validation.

4. SUMMARY

An existing model of human cognition, perception and action (ACT-R) was extended to better support the modeling of human vision in dynamic 3D environments. The extensions provide improved support for motion perception and extend spatial encoding into three dimensions.

In motion perception, the detection and encoding of the direction and magnitude of the motion of objects across the visual field of a digital human model within a COTS virtual environment was implemented. The ability to guide search to motion features was also implemented. This led to the implementation of the ability to encode the features of and recognize moving objects. In addition, the impact of changes in motion on visual attention was implemented based largely on the work of von Mühlenen, et al. (2005).

In spatial encoding, an egocentric representation of the visual-location system of the vision module was implemented. When an object is the focus of attention, the object's egocentric bearing, pitch, and distance relative to the location of the digital human model's head in the virtual environment are added to the features encoded by the visual system. This 3D representation of the object location is used in part to maintain awareness of objects that are no longer visible in the visual field. In our future work to extend the motor capabilities of the ACT-R architecture, these spatial locations will be used to drive the movement of end effectors such as the hands or the feet to locations to interact with objects.

The addition of these capabilities to the ACT-R cognitive architecture allows cognitive models to *see* visual percepts in dynamic 3D virtual environments developed in the COTS software package, Virtools. The next step is to validate these extensions by developing a model of a simple visual task using each of the extensions and directly comparing the quantitative data generated by ACT-R to data collected from human participants. After the model has been validated, attempts can be made to use the model for evaluating real-world tasks relevant to FCS or the FFW system. Future work with ACT-R will also include extensions of the motor system to support modeling human interaction with object prototypes within the environment.

ACKNOWLEDGEMENTS

This research was conducted in the Human Factors and Ergonomics Lab at the Center for Advanced Vehicular Systems at Mississippi State University. Funding for this research was provided as part of a grant from the US ARMY TARDEC National Automotive Center.

REFERENCES

- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y., 2004: An integrated theory of the mind. *Psychological Review*, **111**, 1036-1060.
- Anderson, J. R., Matessa, M., & Lebiere, C., 1998: The visual interface. *The Atomic Components of Thought*, J.R. Anderson and C. Lebiere, Eds., Erlbaum, 143-166.
- Ball, J. T., Gluck, K. A., Krusmark, M. A., and Rodgers, S. M., 2003: Comparing three variants of a computational process model of basic aircraft maneuvering. *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*, Orlando, FL, Institute for Simulation and Training, 87-98.
- Driver, J., McLeod, P., and Dienes, Z., 1992: Are direction and speed coded independently by the visual system? Evidence from visual search. *Spatial Vision*, **6**, 133-147.
- Gunzelmann, G., and Anderson, J. R., 2006: Location matters: Why target location impacts performance in orientation tasks. *Memory & Cognition*, **34**, 41-59.
- Gunzelmann, G., Anderson, J. R., and Douglass, S., 2004: Orientation Tasks Involving Multiple Views of Space: Strategies and Performance. *Spatial Cognition and Computation*, **4**, 209-256.
- Harrison, A. M., and Schunn, C. D., 2003: ACT-R/S: Look Ma, no "cognitive map"! *Proceedings of the Fifth International Conference on Cognitive Modeling*, Bamberg, Germany, Universitäts-Verlag Bamberg, 129-134.
- Hiatt, L. M., Trafton, J. G., Harrison, A. M., and Schultz, A. C., 2004: A cognitive model for spatial perspective taking. *Proceedings of the Sixth International Conference on Cognitive Modeling*, Pittsburgh, PA, 354-355.
- Hillstrom, A.P., and Yantis, S., 1994: Visual motion and attentional capture. *Perception & Psychophysics*, **55**, 399-411.
- Johnson, T. R., Wang, H., and Zhang, J., 2003: An ACT-R model of human object-location memory. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, MA, 1361.
- Liu, Y., Feyen, R., and Tsimhoni, O., 2006: Queueing network-model human processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction*, **13**, 37-70.
- Lyon, D. R., Gunzelmann, G., and Gluck, K. A., 2006: Key components of spatial visualization capacity. *Proceedings of the Seventh International Conference on Cognitive Modeling*, Trieste, Italy, 381-382.
- McLeod, P., Driver, J., and Crisp, J., 1988: Visual search for conjunctions of movement and form is parallel. *Nature*, **322**, 154-155.
- Porter, J. M., Case, K., and Freer, M. T., 1999: Computer-aided design and human models. *Handbook of Occupational Ergonomics*, W. Karwowski and W. Marras, Eds., CRC Press LLC, 479-500.
- Posner, M. I., 1980: Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3-25.
- Royden, C. S., Wolfe, J. M., and Klempen, N., 2001: Visual search asymmetries in motion and optic flow fields. *Perception & Psychophysics*, **63**, 436-444.
- Salvucci, D. D., 2006: Modeling driver behavior in a cognitive architecture. *Human Factors*, **48**, 362-380.
- Salvucci, D. D., 2001: An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, **1**, 201-220.
- Thomas, M. D., Carruth, D. W., McGinley, J. A., and Follett, F., 2006: Task irrelevant scene perception and memory during human bipedal navigation in a genuine environment. *Proceedings of the 25th Army Science Conference*, Orlando, FL.
- Triesman, A. M., and Gelade, G., 1980: A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- Tversky, B., 2003: Structures of mental spaces: How people think about space. *Environment and Behavior*, **35**, 66-80.
- Tversky, B., Morrison, J. B., Franklin, N., and Bryant, D. J., 1999: Three spaces of cognition. *Professional Geographer*, **51**, 516-524.
- Shelton, A. L., and McNamara, T. P., 2001: Systems of spatial reference in human memory. *Cognitive Psychology*, **43**, 274-301.
- von Mühlenen, A., Rempel, M. I., and Enns, J. T., 2005: Unique temporal change is the key to attentional capture. *Psychological Science*, **16**, 979-986.
- Waller, D., Montello, D. R., Richardson, A. E. & Hegarty, M., 2002: Orientation specificity and spatial updating of memories for layouts. *Journal of Experimental Psychology, Learning, Memory & Cognition*, **28**, 1051-1063.
- Wang, R. F., and Spelke, E. S., 2000: Updating egocentric representations in human navigation. *Cognition*, **77**, 215-250.
- Wolfe, J. M., 1998: Visual Search. *Attention*, H. Pashler, Ed., University College London Press, 13-74.
- Wolfe, J. M., 1994: Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, **1**, 202-238.